Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
CrossMark
Procedia - Social and Behavioral Sciences 140 (2014) 447 – 451
Procedia
Social and Behavioral Sciences

PSYSOC 2013

Exploring the Usefulness of a Decision Tree in Predicting People's Locations

Jae Sung Lee [a] & Eun Sung Lee [b]*

[a] Department of Management Sciences, London School of Economics and Political Science
London WC2A 2AE, United Kingdom
[b] School of Business, Yonsei University
Yonsei-Ro 50, Seodaemun-Gu, Seoul 120-749, Republic of Korea

Abstract

Location prediction has been investigated by many researchers. However, there are few studies that explore the relationship between human activities and location. This paper proposes the use of a decision tree to investigate how location can be predicted after observing a series of human activities. A decision tree provides a set of easy-to-interpret decision rules that are necessary for decision makers to be able to make timely and appropriate decisions about location prediction. Based on more than 6000 contextual datasets obtained from college students, we conducted experiments with the WEKA software. Our findings revealed that, given a number of human activities and personal information, a decision tree classifier provides a set of useful rules through which appropriate inferences about location prediction are possible.

Selection and peer-review under responsibility of the Organizing Committee of PSYSOC 2013.

Keywords: Location Prediction, Human Activities, Decision Tree, Context Prediction

Introduction

Predicting the movements of people is important, particularly in the domains of social science and ubiquitous decision support, among others. Regarding the prediction of people's movements, this paper attempted to predict which activity a person would engage in after having performed another specific activity. The logic is based on the assumption that there are repetitive patterns in human's movements (Gonzalez, et al., 2008). The type of human

* Corresponding author: Eun Sung Lee.
E-mail address: starystar725@gmail.com

Selection and peer-review under responsibility of the Organizing Committee of PSYSOC 2013.
doi:10.1016/j.sbspro.2014.04.451

movement this paper is interested in especially is location prediction given a series of human activities. To the best of our knowledge, this has rarely been explored in the literature. For this purpose, we proposed the use of a decision tree to extract a set of relevant rules from a contextual dataset.

A contextual dataset is composed basically of user contexts that explain a number of situational attributes in which the users are performing specific types of activities at specific moments in time (Lee and Cho, 2012). Proper location prediction must be achieved through sophisticated analyses of the contextual dataset before providing relevant location-based services (Vu et al., 2009). However, in this era of rampant and ubiquitous computing based on highly sophisticated mobile devices, such as smart phones, tablet computers, and laptops, naïve location prediction, which has been conducted extensively in the literature to date no longer offers significant implications. Rather, it becomes more important for telecommunications companies (or similar decision makers) to predict reliably where people, who has specific personal information and does some activities, will arrive.

For this purpose, our paper proposes the use of a decision tree to provide a set of meaningful rules that show a series of people's activities through which proper location prediction is finally possible. Experimental results from a contextual dataset collected from college students revealed that enhanced proper location prediction is possible when personal information and their activities are known.

2. Previous Studies

2.1. Location Prediction

The main theme of this paper deals with location prediction given people's previous activities, which is rare in the literature. Before conducting such research, let us briefly review recent trends in location prediction. This topic has drawn much attention from both practitioners and researchers (Yava et al., 2005; Vu et al., 2009; Do & Gatica-Perez, 2013). However, all the previous studies have some pitfalls. First, most previous studies have assumed the existence of user movement patterns, which should have been discovered beforehand to increase the practicability of location prediction. Second, they require knowledge of the probability distribution of moving people's velocity and/or directions of movement (Rajagopal et al., 2002). Moreover, a certain level of noise in moving people's mobile information has triggered significant distortion in the quality of location prediction.

To overcome these pitfalls, we propose that moving users' previous activities may be taken into consideration before making proper location prediction. This approach has a number of noteworthy advantages. First, location prediction provides rich information about the reasons why users are travelling to a particular location. However, location prediction results, combined with users' previous activities, offer rich clues in understanding their hidden intentions, through which more suitable and timely services (or contents) may be supplied to them. Second, users' previous activities and other related information can be identified by using simple if-then grammars. Thereafter, users' intentions to visit a target location can be analyzed clearly so that proper services or contents can be proposed. In this way, we believe that our approach will enhance users' satisfaction with ubiquitous decision support mechanisms available on various types of mobile devices, such as smart phones, tablet PCs, and laptops.

2.2 Decision Trees

The C4.5 decision tree algorithm was adopted as the primary vehicle for executing a decision tree for our experiment. Decision trees are very popular in deriving if-then rules that are human-readable and easy to use in practical applications. Basically, the C4.5 algorithm is capable of approximating discrete-valued functions represented as a decision tree where there are root nodes, child nodes, branches, and leaf nodes (Quinlan, 1993, 1996). Target nodes are usually represented as leaf nodes. The C4.5 decision tree learning mechanism is based on a heuristic, one-step look ahead, and non-backtracking search through the space of all possible decision trees (Polat and Gunes, 2009). Its main steps are as follows:

- Select an attribute and formulate a logical test of attributes
- Branch on each outcome of the test, and move to the corresponding child node the subset of examples (training data) that satisfy that outcome

- Run recursively on each child node
- Termination rule specifies when to declare a leaf node to prevent over-fitting

3. Experiment

3.1 Contextual Dataset

In order to collect the dataset, we asked students to fill out Excel templates that described their travel routes around the campus for a period of two weeks, as well as various types of activities in which they engaged during that time. They were shown a table of current activity codes as shown in Table 1 and a map containing building and route information. In addition, the test subjects filled out a form that asked for basic information, such as gender, major, weekday leisure activities, and lunch time leisure activities. A total of 335 students participated in this survey. As a result, we collected a total of over 6000 datasets with 20 attributes: "day of week;" "prior activity;" "location;" "current activity;" "gender;" "major;" "grade;" "military service;" "religion;" "smoker;" "boy or girlfriend;" "weekday leisure activity;" "weekend leisure activity;" "activity during lunch time;" "level of use of leisure time;" "level of satisfaction with leisure activity;" "housing;" "transportation coming to school;" "transportation coming home from school," and "average study time per day". These data were used to construct a decision tree model.

Table 1. Locations and User Activities

(a) Name of Locations (23 building in total)

Suseon Hall	CentralLibrary	InternationalHall
BusinessBuilding	GeumjandiSquare	OutsideCampus
EastGate	HoamHall	RearGate
StudentUnion	FrontGate	600thAnniversaryBuilding
DasanHallofEconomics	ToegyHallOfHumanities	FacultyHall
LawBuilding	BasketballCourt	SuseonHallAnnex
LargePlayground	ShuttleBusStop	Tanghyeongwan
Bicheondang	Yurimhoegwan	

(b) User Activities

Activity Code	Corresponding Activity
01	Lecture
02	Preparation for Exams/Projects
03	Other Academic Works
04	School Jobs
05	Social Activity(Including Student Union)
06	Chatting
07	Getting Counseling
08	Shopping
09	Having Snacks/Refreshments(Including Smoking)
10	Having a Meal
11	Financial Transaction(Including Posts)
12	Exercising
13	Casual Activities/Games
14	Using Services (eg. Photocopier, Hair Salon)
15	Job Hunting
16	Internet Surfing
17	Other

With the aid of Weka [Hall et al., 2009] which is available at http://www.cs.waikato.ac.nz/ml/weka/, the 20-variable dataset was used to determine networks with "Location" as a target node. The structure of the decision tree was constructed using the J48 algorithm, which is powered by C4.5.

3.2 Results

Let us check the rules generated by a decision tree for the Business Building and Law Building, which are known as two major locations for the college students in the private university where one of the authors works. True positive rate (or TPR) for the Business Building was 0.649, and for the Law Building was 0.806. Selected decision tree results for the two locations were as follows.

Rule 1:
IF (Activity Code = 5, 6, 7, 8, 9) and (Grade = Freshman) and (Smoking = Yes) and (Major = Free Major, Confucianism and Eastern Philosophy, Liberal Arts, Law) THEN (Location = Law Building).

Rule 2:
IF (Activity Code = 2, 3, 4) and (Grade = Sophomore, Junior, Senior) and (Level of using leisure time = Very Low) and (Major = Free Major, Confucianism and Eastern Philosophy, Liberal Arts, Law) THEN (Location = Business Building).

Rule 3:
IF (Activity Code = 3, 4) and (Grade = Junior, Senior) and (Level of using leisure time = Not Very Low) and (Level of satisfaction with leisure activity = Positive to Very Positive) and (Major = Business, Domestic Science) and (Activity during lunch time = Team sports, Recreation, Viewing, Socialization, Walking) and (Average study time per day = More than 2 hours) and (Transportation for coming home from school = Walk, Private Car, (Auto) Bicycle) THEN (Location = Business Building).

As shown above, interpreting these rules is very simple and easy for users. Let us elaborate on the interpretation of the three rules above. First, depending on specific location, users seem to show different activities according to different personal information. For example, from the activity codes in Rule 1, we learned that when students are relatively relaxed, they seem to visit the Law Building. In contrast, we learned that students go to the Business Building when they have relatively serious things to do, such as a study seminar and/or administrative needs, which is also easily determined from interpreting the activity codes in Rules 2 and 3. Second, from Rule 3, we discovered that when students majoring in business and domestic science go to the Business Building, they are quite active in various activities compared to the students with other majors, such as Confucianism, liberal arts, and law. However, this rule works only when there is no exam (activity code 2).

4. Concluding Remarks

In this paper, we used a decision tree model to perform location prediction based on a contextual dataset collected from college students. Overall performance was that TPR was 0.51 and ROC area was 0.837. Considering the 24 locations to be classified correctly, the performance of the decision tree model like this is promising. One of the advantages we found in using the decision tree model for the task of location prediction was that it produces a set of rules that are very easy to interpret and therefore are simple to apply to real ubiquitous decision support applications. This advantage will serve as a critical factor in deciding which method to use to support practical context prediction, as it provides a clear view of the interrelationships between numerous variables and can handle large and noisy datasets. In future studies, we intend to analyze the same problem with other related techniques, such as Bayesian and neural networks, in order to compare their performance to the decision tree model.

References

Do, T.M.T. & Gatica-Perez, D. (2013). Where and what: Using smartphones to predict next locations and applications in daily life. *Pervasive and Mobile Computing*. Retrieved from http://dx.doi.org/10.1016/j.pmcj.2013.03.006.

Gonzalez, M.C, Hidalgo, C.A., & Barabasi, A.L. (2008). Understanding individual human mobility pattern. *Nature,* 453, 779–782.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11, 10–18.

Quinlan, J. R. (1993).*C4.5: Programs for machine learning*. Kluwer Academic Publishers.

Quinlan, J. R. (1996). Improved use of continuous attributes in C45. *Artificial Intelligence*, 4, 77–90.

Lee, K.C. & Cho, H. (2012). Integration of General Bayesian Network and ubiquitous decision support to provide context prediction capability. *Expert Systems with Applications*, 39, 6116-6121.

Polat, K. & Gunes, S. (2009). A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. *Expert Systems with Applications*, 36, 1587-1592.

Rajagopal, S., Srinivasan, R. B., Narayan, R. B., & Petit, X. B. C. (2002). GPS-based predictive resource allocation in cellular networks. *Proceedings of the IEEE International Conference on Networks*, 229–234.

Vu, T.H.N., Ryu, K.H., & Park, N. (2009). A method for predicting future location of mobile user for location-based services system. *Computers & Industrial Engineering*, 57, 91-105.

Yava, G., Katsaros, D., Ulusoy, O., & Manolopoulos, Y. (2005). A data mining approach for location prediction in mobile environments. *Data & Knowledge Engineering*, 121–146